Plan to Measure SAM Limits

Adam Lyon and the SAMGrid Team

1 Motivation

The recent release of CDF SAM requirements leads one to ask the question, "What are the maximum rates that SAM data handling can achieve?" The answers for a CDF production like environment are unknown. This document proposes a short plan of tests to determine SAM services levels using a CDF CAF environment.

Tests designed here can also be used as validation tests for future SAM releases. Such tests would ensure that new releases continue to meet the agreed upon SAM service levels.

2 Methodology

We will perform tests using the CDF test CAF (the phase one farm)[1] configured to 1000 virtual machines. For each test, two DBServers are used -- one for the CAF station and another for the CAF submission and worker clients. The integration universe is to be used for all tests. The SAM test harness is utilized to execute CAF submissions in a controlled and ramping fashion.

Each test must include the following information

- Statement of the test object
- Prescription for running the test
- Description of monitoring and instrumentation to collect data during the test
- If the test is for validation, then a definition of passing
- A description of a write up that documents the test

3 The Tests

The following sections detail the tests that we wish to accomplish.

---

[1] The SAM Team thanks CDF for allowing us use of their phase 1 farm for these and future tests.

3.1 File Delivery Limit Tests

3.1.1 Test object

Determine the limit of the file delivery rate in a CAF like environment. The result is the maximum number of file deliveries per minute that SAM can attain with 1000 simultaneous virtual machines accessing the same SAM station, the same station DB server, and the same client DB server.

3.1.2 Prescription

There are two "axes" for these tests. One is the *get next file request rate*. This value is indirectly controlled by a sleep time in the CAF job. The other axis is the degree of data set file overlap. The latter is important because the station behaves differently if all jobs are asking for the same files than if all jobs are asking for different files. We plan to test both extremes (all datasets are made up of the same files and none of the data sets have files in common). We also plan to test several values of the sleep time (perhaps 1 minute, 30 seconds, 15 seconds, 1 second -- we'll have to see what the system can handle).

The test CAF is started with 100 jobs of 10 segments each. Each segment will ask for 10 files (therefore each dataset must contain 100 files). The sleep time of the job dictates the get next file request rate (e.g. the job script merely loops over issuing a get next file, sleeping for a period of time, and then issuing a release file).

Jobs are continually submitted to the CAF so that all 1000 VMs are active (the submission interval must be worked out). The test should run at a maximal get next file rate for thirty minutes. After that time, all CAF jobs are killed.

3.1.3 Measurements

We wish to measure the following:

- The *get next file* rate. This is measured by inserting timing information into the CAF job so that the wait time is printed to the job log file.
- The file delivery rate. This can be measured by parsing the SAM station log.
- The file delivery error rate. This can be measured by parsing the SAM station log.
- The station CPU and memory usage. This can be obtained with an external script from Jerry.
- The station DB server CPU and memory usage. This can be obtained with an external script from Jerry.

- The client DB server CPU and memory usage. This can be obtained with an external script from Jerry.
- The number of servant connections for each DB server. This can be obtained from the DB server logs.
- The DB server query response time. This can be obtained from the DB server logs.
- The amount of time it takes for the "sam submit project" or equivalent command to run.

3.1.4 Documentation

A document is to be produced with plots of the measurements specified in the previous section vs. time. For each sleep wait time value and file overlap value, the mean and maximal values of the measurements are documented.

3.2 File meta-data declaration tests

3.3 File meta-data read tests

3.4 File meta-data undeclared tests

3.5 Simple Farm file store operation with a background of a busy test CAF